DOCUMENT RESUME

ED 263 767 FL 015 388

AUTHOR Browning, Gari
TITLE What Makes ESL Students' Speech Sound Unacceptable?
PUB DATE 85
NOTE 11p.; Paper presented at the annual meeting of the California Association of Teachers of English to Speakers of Other Languages (16th, San Diego, CA, April 1985). In: CATESOL (California Association of Teachers of English to Speakers of Other Languages) Occasional Papers, Number 11, Fall 1985. p72-81.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *English (Second Language); *Error Patterns; Foreign Students; Interpersonal Communication; Negative Attitudes; *Peer Acceptance; *Phonology; Second Language Instruction; *Speech Habits; Work Environment

ABSTRACT

A study of the gravity of non-native speakers' speech errors, particularly as viewed in the workplace, was based on two assumptions: that certain features of spoken English contribute more to speech acceptability than others, and that native speakers have an internalized, ordered list of criteria for making judgments about non-native speakers' English. The study had three groups of subjects: 70 students of English as a second language (ESL) who provided speech samples; 29 ESL teachers who judged the samples' acceptability, identified non-native features using a checklist, and ranked the five least acceptable features; and 21 naive judges chosen on the basis of their occupations, with little linguistic or foreign language training and limited exposure to ESL, who judged which of a pair of artificial speech samples was more acceptable. While results suggest a hierarchy of speech sound acceptability generalizable on the basis of occupation, further research on isolating features and other independent variables is recommended. Although some variation from earlier hierarchies was found, the methodology used was determined to be viable for extending the hierarchy. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

ED263767

California Association
Teachers of English to
Speakers of Other Languages

CATESOL Occasional Papers
Number 11 (Fall, 1985)

WHAT MAKES ESL STUDENTS' SPEECH SOUND UNACCEPTABLE?

Gari Browning
Orange Coast College, Costa Mesa

The most influential judges of the English of ESL students are outside the classroom--on the job. This paper identifies the most important features of oral language from the listener's perspective. Using recorded speech samples, a group of listeners involved in job interviewing chose the features which made each sample sound unacceptable. A list of features important for the non-native English speaker resulted. This list has broad implications for teaching ESL. Knowing what the real world considers unacceptable in ESL speech can help the teacher prioritize class time and decide which are the most crucial errors in students' speech.

Many of us have seen second language learners, maybe even our own students, fail in their attempts at oral communication outside class, even students considered to be able to communicate well in the class. This dilemma is particularly evident in situations where the ability to communicate in the second language is crucial. An example situation is the job interview in which many non-native speakers of English are turned away because their control of the language is judged inadequate. "Judged" is a key word here because these interviewers are judging the English of the applicant. What kind of judges do they make? What is it about the non-native speaker's speech that makes it sound unacceptable to these judges?

The study designed to address these questions is an error gravity study as well as an attitude study. It's an error gravity study because it puts errors in order of seriousness, an attitude study since it uses opinions or attitudes of judges as

---

This paper was presented at the 1985 CATESOL Conference held in San Diego, California.

Gari Browning has taught ESL since 1968, both in the United States and abroad and has recently taught in the TESL M.A. program at the University of Hawaii at Manoa. She holds a Ph.D. in Applied Linguistics from UCLA and has done research in language testing and speaking. Currently, she is the director of the American Language Program at Orange Coast College.

FL015388

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

72

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert Gilmore

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

the criteria for ordering the errors.

Attitude studies have shown negative reactions on the part of personnel officers toward non-standard dialects. Shuy (1970) showed that, on the basis of speech alone, members of minority groups are perceived unfavorably by prospective employers. The possibility that similar negative reactions to speech extend to non-native speakers of English cannot be ignored by language educators. What can we do? Error evaluation based on subjective reactions can provide insights into the sources of such negative reactions. Specifically, it can tell us what linguistic features trigger negative reactions. By knowing what linguistic factors contribute to these reactions, we may get a better idea of when students have mastered those aspects of oral English that would otherwise contribute unfavorably to their job interviews or other important communicative situations.

The study was undertaken to determine the impact of different kinds of linguistic errors on the effectiveness of oral communication. How can we characterize the "effectiveness of oral communication"? For the purpose of this study we will define it as acceptable speech. Acceptability is defined by Chomsky (1965) as that which is natural, comprehensible, and not bizarre or outlandish in any way. The basic elements of this definition are comprehensibility and irritation. Unusualness was equated with irritation in this study. Irritation is presumed to be the reaction to unnatural, bizarre, or outlandish utterances. Johansson (1979) based his grading of errors on comprehensibility and irritation. The definition of acceptable/unacceptable used in this study closely parallels Johansson's error gravity system. An adaptation of his model clearly shows the relationship of comprehensibility and irritation to acceptability and unacceptability. Figure 1 shows how a determination of acceptability is made for a speech feature which is not native. Looking at Figure 1, we see that if no non-native feature is present, the segment will obviously be judged linguistically acceptable, at least for our purposes. If a non-native feature is present, it may or may not be noticed. If it is not noticed, there may still be some undefinable negative reaction to the speech; there may be irritation. If there is irritation, then the feature will be judged unacceptable. If there is no notice of the feature and no irritation, the feature will be passed over as acceptable. If the feature is noticed, the listener first attends to the meaning of the utterance. If the meaning is comprehended, the feature is judged as unacceptable since it has affected communication in a broader sense.

In applying this model, two basic assumptions were made: 1) certain features of spoken English contribute more to acceptability of a sample of speech than others, and 2) native speakers have an internalized, ordered list of criteria for making judgments about ESL speech. Given the model of what happens when acceptability judgments are made, and assuming there

Figure 1
The Relationship of Comprehensibility and Acceptability

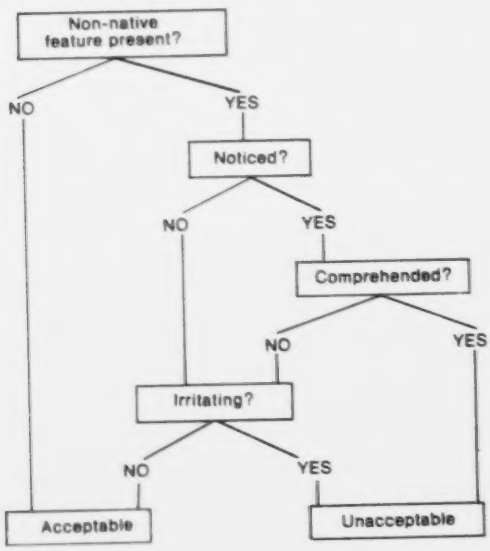
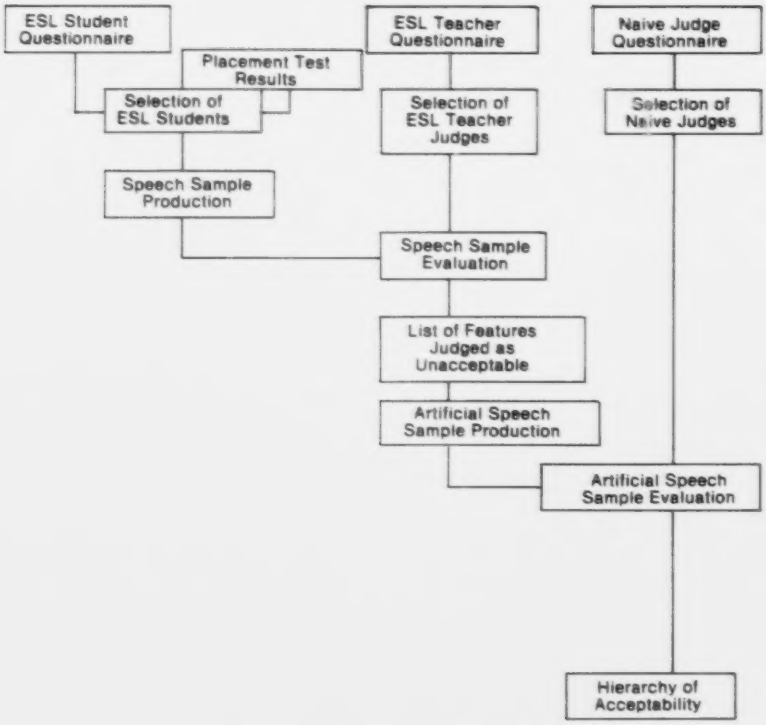


Figure 2
Research Design Chart



is an underlying order of acceptability to non-native speech features, how can the order or hierarchy be discovered? This question brings us to the rather complicated design of the study.

As illustrated in Figure 2, first, three groups of subjects were needed. There was a group of ESL students, a group of ESL teachers, and a group of naive judges. The group of ESL students numbered 70. They recorded one-minute speech samples which were their descriptions of a selected picture. Based on the first language of the speaker, his/her ESL level (as determined by an independently administered test), and the grammatical level of the sample, 19 non-native speech samples giving the most variety possible were selected for judging by the group of ESL teachers.

There were 29 ESL teachers selected on the basis of a combination of ESL teaching experience and training. They had three things to do with the non-native speech samples: 1) they judged them each globally on a 4-point scale of acceptability; 2) they identified all of the non-native features they noticed on each sample using a checklist (see Table 1); and 3) they then ordered the five most unacceptable features for each sample. (A later check comparing the features the teachers identified and the features revealed by careful analysis showed that the teachers tended to identify features accurately, misidentifying only 7%. They tended to miss some of the features, however. (The ability to identify features on the checklist seemed to be related to training, perhaps more specifically to familiarity of the terms on the checklist.) The purpose of this step was to show which features might be most profitably included on a second set of speech samples to be judged by naïve judges.

There were 21 naïve judges selected on the basis of occupation, lack of linguistic and foreign language training, and limited exposure to ESL. These naïve judges judged a second set of speech samples which were presented in pairs. Each sample was artificially produced and contained a feature or combination of features. Each was judged simply on the basis of which of a pair was more acceptable. Based on these results from the naïve judges an ordered list of features was established.

Along the way some interesting things were discovered, especially during the preparation of the checklist the ESL teachers used. Specifically, some comparisons were made between the speech samples made by the 70 ESL students and some made by native English speakers during a pilot study. The task for both groups was the same, but the analyses showed striking differences. For example, on a one-minute speech sample non-native speakers used an average of 2.3 different tenses, while native speakers used 6.1 tenses. Non-natives used 14% complex sentences compared to the natives' 39%. The ratio of content to function words used was very similar, with the native speakers using only slightly more, but since the task was

Table 1

Scoring of Non-Native Speech Features by ESL Teachers

Checklist of Non-Native Features	Average Score on 6-Point Scale	Total Votes
Pronunciation		
1. mispronounced stressed vowels	3.00	184
2. mispronounced unstressed vowels	3.12	108
3. mispronounced single consonants	2.79	226
4. mispronounced consonant clusters	3.07	175
5. inappropriate intonation	2.09* (3)	176
6. misplaced word stress	2.46* (2)	112
Grammar		
7. lack of grammatical complexity	3.41	127
8. lack of grammatical variety	3.82	137
9. lack of global grammatical correctness	2.58* (1)	141
10. lack of local grammatical correctness	1.92* (4)	332
Vocabulary		
11. lack of variety in vocabulary	3.53	76
12. lack of correctness in vocabulary	2.98	184
Fluency		
13. short filled pauses	3.75	138
14. long unfilled pauses	2.85	108
15. inappropriate unfilled pauses	3.26	114
16. non-English phenomena	3.41	91
17. false starts	3.39	80

*most unacceptable features; features used in artificial speech samples; numbered from most to least acceptable

"describing a picture," the content vocabulary needed may have been limited. The total number of words used differed greatly between the two groups, however, and had a lot to do with some of the statistics just reported. The average number of words on the

one-minute sample for the native speakers was 307.7, and for non-native speakers, 85.25. This shows that the native speakers spoke much faster; that is, they had fewer and shorter pauses and hesitations. According to Maclay and Osgood (1959) and Crystal and Quirk (1964), speed is a function of how much the speaker hesitates and not of how quickly he pronounces the words.

The second set of speech samples was used in the final phase of the study. As a result of the ESL teacher judgments of features, the list of 17 features was reduced to four for the sake of manageability. The four are marked with an asterisk on Table 1. They are: inappropriate intonation, misplaced word stress, lack of global grammatical correctness, and lack of local grammatical correctness. The global/local distinction is from Burt and Kiparsky (1974). Global errors are those which violate the rules involving the overall structure of the sentence and relations among clauses or the major parts of the sentence, S-V-O, for example, "My friend said that not take this bus, we are late for school." Local errors are confined to one part of the sentence, for example, "I should like know...." The four features mentioned were used in making the artificial set of speech samples.

Once selected, all of the possible combinations of the four features were determined. Sixteen combinations resulted. A native speaker then recorded a 20 to 30-second segment of speech for each combination of features plus one which was feature-free (strictly native English). All samples were completely native English except for the combination of features to be represented. In order to make these artificial samples as representative of the non-native speech as possible, the frequency and distribution of the non-native features were based on the features as they occurred in the original one-minute non-native speech samples. Also, actual errors as they had occurred in the non-native samples were used to ensure the authenticity of the features. Therefore, the frequency, distribution, and actual errors were very close to real non-native speech.

The artificial speech samples were then paired for easy comparison by the 21 naïve judges. As mentioned, these naïve judges had limited language exposure and no training, but they were also chosen on the basis of their occupations; ten were personnel officers. More judges were added from other occupations the ESL student might encounter. They were: 4 health service workers, 2 policemen, 2 attorneys, 1 welfare worker, 1 postal clerk, and 1 bus driver. The naïve judges did three things: they rated the original 19 non-native speech samples using the same four-point scale the ESL teachers had used. Then they judged the artificial speech samples by listening to a pair and choosing the one they found more acceptable. Then using a 5-point Likert scale, they decided how much more acceptable the one they chose was. This task was arranged so that the naïve judge would not have to deal with

linguistic terms and not even need to identify what was causing his or her negative reaction to the speech sample.

In order to interpret these judgments on the Likert scale, I devised average order-ratings for each feature. Then I compared them using an Analysis of Variance and a Tukey test of honest significant difference. These analyses showed that there was enough difference in the judgments on the Likert scale to yield a hierarchy of acceptability for those four features, i.e., there was an order. The hierarchy is in Table 2. Regardless of whether the features appeared with other features or in isolation, the order remained the same.

Table 2

Hierarchy of Acceptability for Four Features of Non-Native English (From All Artificial Speech Samples)

Feature	Average Order-Rating (derived from points assigned in ordering portion of teachers' rating task).	
Lack of global grammatical correctness	1.3245	most acceptable
Inappropriate intonation	1.0948	
Misplaced word stress	0.9779	
Lack of local grammatical correctness	0.8025	least acceptable
	($F = 13.09^*$, $p < .05$)	

The only violation of this order was in the judgments made by the ESL teachers. Their order from most acceptable to least acceptable was: global grammatical errors, word stress, intonation, and local grammatical errors, with the two middle features reversed. However, the ordering of these four was not important at that point in the study since the purpose of the teachers' task was to select features for the naïve judgments. The reader will recall that the teachers had 17 features to choose from and the features were not in isolation. More importantly, there was no significant difference in the teachers' judgments or the naïve judgments for word stress and intonation, suggesting that the ordering of these two features is less valid than that for the other features.

Conclusions

First, we may conclude that there is a hierarchy of acceptability. It may be generalizable on the basis of occupation, but we don't know about other factors such as educational level and dialect (participant factors). Nor do we know if the hierarchy would hold up in a different second language situation.

Second, co-occurrence of features doesn't make a difference; the hierarchy remains the same. This suggests that further investigation could use isolated features.

Next, it should be noted that a couple of sets of features which were ordered in previous studies did not have the same ordering in this study. The first set of features that did not conform to previous research findings was Burt and Kiparsky's global and local grammatical errors. Local errors were found to be more unacceptable than global errors in my study. It is unclear in the Burt and Kiparsky study exactly how the judges were exposed to the sample, i.e., whether it was silently read, read aloud to them, or played on a tape using the original non-native speaker's voice. Methodology may, therefore, account for the difference.

The other set of features ordered differently in this study from other studies was vocabulary errors and grammar errors. Results of studies by Johansson (1978), Engh (1971) and Politzer (1978) all conclude that vocabulary errors are more unacceptable than grammatical errors. Some differences between my study and their studies should be pointed out, however. First, both Johansson and Engh studied errors in writing rather than speaking. Although Politzer studied speech, he used German, not English. The most important difference was in methodology: in all three studies mentioned, grammatical and vocabulary errors were compared one to one. But in actual speech the occurrence of these errors is not equal. I found that grammatical errors occurred about twice as often as vocabulary errors. From this evidence we can conclude that frequency of occurrence is indeed a factor in deciding the acceptability of a feature.

Finally, the methodology used in this study is viable and can be used to extend the hierarchy. Once complete it would provide insight that may enable us to better assist our students before they leave their ESL training.

References

Burt, Marina and Carol Kiparsky. 1974. Global and local mistakes. In New frontiers in second language learning, J. H. Schumann and N. Stenson (Eds.), Rowley, Massachusetts: Newbury House.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge: Massachusetts Institute of Technology.

Crystal, D. and R. Quirk. 1964. Systems of prosodic and paralinguistic features in English. The Hague: Mouton.

Engh, B. 1978. En toleransundersökning: tyska elever tolkar svenska elevers språkfel. Pedogogisk-psykologiska problem. Lärarhögskolan, Malmö. Reviewed in Studies of error gravity, S. Johansson, Guthenburg: Acta Universitatis Gothoburgensis.

Johannson, Stig. 1978. Studies of error gravity: native reactions to errors produced by Swedish learners of English. Guthenburg: Acta Universitatis Gothoburgensis.

Johansson, Stig. 1979. The study of error gravity: a contribution to language testing? Manuscript, University of Oslo, Norway.

Maclay, H. and C. E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. Word (15).

Politzer, R. 1978. Errors of English speakers of German as perceived and evaluated by German natives. Modern Language Journal (62).

Shuy, Roger. 1970. Employee selection, training, promotion: pitfalls of good intentions. Paper presented at C.A.L. - N.C.T.E. Conference of Education and Training in the National Interest: The Role of Language Variety.